

ENERGY OPTIMIZATION FOR DISTRIBUTIONS ON THE SPHERE AND IMPROVEMENT TO THE WELCH BOUNDS

YAN SHUO TAN

ABSTRACT. For any Borel probability measure on \mathbb{R}^n , we may define a family of eccentricity tensors. This new notion, together with a tensorization trick, allows us to prove an energy minimization property for rotationally invariant probability measures. We use this theory to give a new proof of the Welch bounds, and to improve upon them for collections of real vectors. In addition, we are able to give elementary proofs for two theorems characterizing probability measures optimizing one-parameter families of energy integrals on the sphere. We are also able to explain why a phase transition occurs for optimizers of these two families.

1. INTRODUCTION

Amongst all Borel probability measures μ on \mathbb{R}^n having the same radial distribution, we seek a minimizer for the energy integral

$$I_k(\mu) := \int_{\mathbb{R}^n} \int_{\mathbb{R}^n} \langle x, y \rangle^k d\mu(x) d\mu(y). \quad (1)$$

In this paper, we will introduce a tensorization trick, thereby proving that the integral is minimized by the rotationally invariant measure, μ_{rot} . More precisely, for any integer k , we define the k -th *eccentricity tensor* of a measure μ . The gap between $I_k(\mu)$ and $I_k(\mu_{rot})$ is then given by the squared Euclidean norm of this tensor. Specializing to Borel probability measures on the sphere, we see that (1) is minimized by the uniform measure. Moreover, we may also adapt the proof to obtain an analogous result for the uniform measure on the sphere in \mathbb{C}^n .

These facts have several interesting applications, the first of which concerns the well-known Welch bounds in the compressed sensing literature. Using the complex case of our result, we recover the original Welch bounds, while using the real case, we are able to improve upon them for collections of real vectors. In our opinion, this proof is more illuminating than the existing ones. It shows one view the Welch bounds as saying that the average cross-correlation of signal sets cannot beat that of the uniform distribution.

Next, we are able to obtain new proofs of Björck's theorem from the 1950s and the recent theorem by Bilyk-Dai-Matzke. These theorems characterize optimizers of two one-parameter families of energy integrals, and were proved using methods from potential theory and spherical harmonics. Our methods have the benefit of being more elementary. Furthermore, our proof scheme for both theorems is very similar, and sheds light on the phase transition phenomenon discussed in [2]. Indeed, we are able to show why the phase transition occurs, and why it happens for different parameter values for the two families.

The plan of the rest of this paper is as follows. In Section 2, we define the eccentricity tensors and use the tensorization trick to prove the energy minimization property of rotationally invariant measures. In Section 3, we discuss the Welch bounds, show how they may be improved, and present some consequences of this improvement. In Section 4, we show how our results imply the two theorems on energy optimization on the sphere, and discuss their relevance to the phase transition phenomenon.

2. ECCENTRICITY TENSORS AND THE TENSORIZATION TRICK

In this section, we shall introduce the tensorization trick, define eccentricity tensors, and prove that rotationally invariant measures minimize (1). For notational as well as intuition purposes, however, it is more convenient to work with random vectors than with measures. We hence do so for the rest of this paper, being careful to assert the independence of collections of random vectors where necessary.

Date: December 20, 2016.

Partially supported by NSF Grant DMS 1265782 and U.S. Air Force Grant FA9550-14-1-0009.

The tensorization trick is to write the integral (1) as the squared norm of the k -th moment tensor of μ .

Notation 2.1. Let X be a random vector in \mathbb{R}^n . For any positive integer k , let

$$M_X^k := \mathbb{E}X^{\otimes k}$$

denote its k -th moment tensor if all entries are finite.

Recall the following fact from linear algebra. For any positive integer k , we may identify the k -th tensor product $T^k(\mathbb{R}^n) = \mathbb{R}^n \otimes \cdots \otimes \mathbb{R}^n$ with \mathbb{R}^{n^k} by picking as a basis the vectors $\{e_{i_1} \otimes e_{i_2} \otimes \cdots \otimes e_{i_k}\}_{1 \leq i_1, \dots, i_k \leq n}$. With this choice, the Euclidean inner product between any two pure tensors $u_1 \otimes \cdots \otimes u_k$ and $v_1 \otimes \cdots \otimes v_k$ can be written as

$$\langle u_1 \otimes \cdots \otimes u_k, v_1 \otimes \cdots \otimes v_k \rangle = \prod_{i=1}^k \langle u_i, v_i \rangle.$$

In particular, for power tensors $u^{\otimes k}$ and $v^{\otimes k}$, we have the formula

$$\langle u^{\otimes k}, v^{\otimes k} \rangle = \langle u, v \rangle^k. \quad (2)$$

Now let X and Y be two independent random vectors. Equation (2) allows us to rewrite the k -th moment of their inner product as an inner product between their k -th moment tensors. Namely, we have

$$\mathbb{E}\langle X, Y \rangle^k = \mathbb{E}\langle X^{\otimes k}, Y^{\otimes k} \rangle = \langle \mathbb{E}X^{\otimes k}, \mathbb{E}Y^{\otimes k} \rangle = \langle M_X^k, M_Y^k \rangle. \quad (3)$$

For independent copies X, X' of the same random vector having distribution μ , $M_X^k = M_{X'}^k$, so

$$I_k(\mu) = \mathbb{E}\langle X, X' \rangle^k = \|M_X^k\|^2. \quad (4)$$

We next introduce the notion of the rotation symmetrization of a random vector.

Definition 2.2. For any random vector X in \mathbb{R}^n , let X_{rot} denote a random vector that is independent of X , has the same radial distribution as X , and whose distribution is rotationally invariant. We call X_{rot} the *rotational symmetrization* of X .

Comparing the moment tensors of a random vector and those of its rotational symmetrization give rise to what we shall call eccentricity tensors.

Definition 2.3. Let X be a random vector in \mathbb{R}^n with finite moments of all orders. For any positive integer k , define its k -th *eccentricity tensor* to be

$$E_X^k = M_X^k - M_{X_{rot}}^k. \quad (5)$$

Since $X \stackrel{d}{=} X_{rot}$ if and only if X is rotationally invariant, we see that the eccentricity tensors of X are quantitative measures of how far its distribution is from being rotationally invariant. This interpretation is further supported by the following observation.

Lemma 2.4 (Orthogonality). *Let X be a random vector in \mathbb{R}^n with finite moments of all orders. Its eccentricity tensors are orthogonal to the moment tensors of its rotational symmetrization. In other words, for any positive integer k ,*

$$\langle E_X^k, M_{X_{rot}}^k \rangle = 0 \quad (6)$$

and

$$\|M_X^k\|_2^2 = \|M_{X_{rot}}^k\|_2^2 + \|E_X^k\|_2^2. \quad (7)$$

Proof. Let Q be a random orthogonal matrix chosen according to the Haar measure on $O(n)$. For any fixed vector $v \in \mathbb{R}^n$, Qv is uniformly distributed on the sphere of radius $\|v\|_2$, so if Y is any random vector independent of Q , applying Q to Y preserves its radial distribution but makes QY rotationally invariant.

Now choose Q to be independent of X and X_{rot} . Our previous discussion implies that

$$Q^T X \stackrel{d}{=} QX_{rot} \stackrel{d}{=} X_{rot}.$$

We use this to compute

$$\mathbb{E}\langle X, X_{rot} \rangle^k = \mathbb{E}\langle X, QX_{rot} \rangle^k = \mathbb{E}\langle Q^T X, X_{rot} \rangle^k = \mathbb{E}\langle X_{rot}, X_{rot} \rangle^k, \quad (8)$$

where X'_{rot} is an independent copy of X_{rot} . We may then apply identities (3) and (4) to rewrite the above equation as

$$\langle M_X^k, M_{X_{rot}}^k \rangle = \langle M_{X_{rot}}^k, M_{X_{rot}}^k \rangle. \quad (9)$$

Subtracting the right hand side from the left hand side gives (6), from which (7) is an immediate corollary. \square

The minimization of the integral (1) by rotationally invariant measures is then an easy consequence of the previous lemma.

Theorem 2.5. *Let X be a random vector in \mathbb{R}^n with finite moments of all orders. Then*

- a) (Minimization) *If X' is an independent copy of X , and X_{rot}, X'_{rot} are independent copies of its rotational symmetrization, we have*

$$\mathbb{E}\langle X, X' \rangle^k \geq \mathbb{E}\langle X_{rot}, X'_{rot} \rangle^k \quad (10)$$

for any positive integer k .

- b) (Uniqueness) *Furthermore, if equality holds in (10) for all k and we further assume that X has a subexponential distribution¹, then X is rotationally invariant.*

Proof. Using identity (4), we rewrite the first claim as

$$\|M_X^k\|_2^2 \geq \|M_{X_{rot}}^k\|_2^2,$$

and this follows immediately from equation (7).

If equality holds for all positive integers k , then by (7), $E_X^k = 0$ for all k , implying that X and X_{rot} have the same moment tensors of all orders. By our assumption on the distribution of the norm and Lemma 2.6 to come, X and X_{rot} have the same distribution. \square

Lemma 2.6. *Let X be a subexponential random vector in \mathbb{R}^n . Then the distribution of X is determined by its moment tensors.*

Proof. Let $K = \|X\|_{\psi_1}$ denote the subexponential norm of X . We then have the following moment growth condition [7]:

$$\sup_{v \in S^{n-1}} \limsup_{k \rightarrow \infty} \frac{(\mathbb{E}|\langle X, v \rangle|^k)^{1/k}}{k} \lesssim K. \quad (11)$$

In particular, every marginal is integrable, so the characteristic function $\phi_X(v) := \mathbb{E}e^{i\langle X, v \rangle}$ is continuous at $v = 0$. This means that X is determined by its characteristic function [4]. Next, condition (11) implies that for each $v \in S^{n-1}$, the function $t \mapsto \mathbb{E}e^{it\langle X, v \rangle}$ can be written as a power series with coefficients $\frac{\mathbb{E}\langle X, v \rangle^k}{k!}$ [1], so $\phi_X(v)$ is in turn determined by the moments $\mathbb{E}\langle X, v \rangle^k$. Finally, observe that $\mathbb{E}\langle X, v \rangle^k = \langle M_X^k, v^{\otimes k} \rangle$ by (3). \square

Corollary 2.7. *Let θ have the uniform distribution on the sphere S^{n-1} , and let X be any random vector taking values on the sphere. Then*

$$\mathbb{E}\langle X, X' \rangle^k \geq \mathbb{E}\theta_1^k = \frac{1 \cdot 3 \cdots (2k-1)}{n \cdot (n+2) \cdots (n+2k-2)} \quad (12)$$

for any positive integer k . Furthermore if equality holds for all k , X has the uniform distribution.

Proof. The inequality and the characterization statement follows immediately from Theorem 2.5. The computation for $\mathbb{E}\theta_1^k$ is the content of the next lemma. \square

Lemma 2.8 (Moments of spherical marginals). *Let θ be uniformly distributed on the sphere S^{n-1} . Then for any unit vector $v \in S^{n-1}$ and any positive integer k , we have*

$$\mathbb{E}\langle \theta, v \rangle^{2k} = \frac{1 \cdot 3 \cdots (2k-1)}{n \cdot (n+2) \cdots (n+2k-2)} \quad (13)$$

¹This is a weak condition satisfied by most distributions dealt with in practice. For an in-depth discussion on the properties of subexponential distributions, we refer the reader to [7].

Proof. There are several ways to prove this identity. We shall prove this by computing gaussian integrals. Let γ and g denote standard gaussians in 1 dimension and n dimensions respectively. Then using the rotational invariance of g , we have

$$\mathbb{E}\gamma^{2k} = \mathbb{E}\langle g, v \rangle^{2k} = \mathbb{E}\langle \|g\|_2 \theta, v \rangle^{2k} = \mathbb{E}\|g\|_2^{2k} \mathbb{E}\langle \theta, v \rangle^{2k}.$$

Rearranging gives

$$\mathbb{E}\langle \theta, v \rangle^{2k} = \frac{\mathbb{E}\gamma^{2k}}{\mathbb{E}\|g\|_2^{2k}}.$$

We then compute

$$\mathbb{E}\|g\|_2^{2k} = \frac{\omega_n}{(2\pi n)^{n/2}} \int_0^\infty r^{2k} r^{n-1} e^{-r^2/2} dr, \quad (14)$$

where ω_n is the volume of the sphere S^{n-1} . It is well known that

$$\omega_n = \frac{2\pi^{n/2}}{\Gamma(n/2)},$$

while we also have

$$\int_0^\infty r^{2k} r^{n-1} e^{-r^2/2} dr = 2^{n/2+k-1} \Gamma(n/2 + k).$$

Substituting these back into (14) gives

$$\mathbb{E}\|g\|_2^{2k} = 2^k \frac{\Gamma(n/2 + k)}{\Gamma(n/2)} = n \cdot (n+2) \cdots (n+2k-2).$$

This yields the denominator in (13). A similar calculation for $\mathbb{E}\gamma^{2k}$ yields the numerator. \square

3. APPLICATIONS TO DICTIONARY INCOHERENCE AND THE WELCH BOUNDS

Given a collection of m unit vectors $Z = \{z_1, z_2, \dots, z_m\}$ in \mathbb{C}^n , we are often interested in the quantity

$$c_{max} = \max_{j \neq k} |\langle z_j, z_k \rangle|.$$

If we think of the vectors as dictionary elements, then c_{max} measures the incoherence or maximum cross-correlation of the dictionary. It is well known in the compressed sensing literature that the larger the value of c_{max} , the worse the collection Z performs when we try to recover a sparse representation of a vector as a linear combination of the z_j 's [6]. As such, it is an important question in the design of communication systems to know how well we can do theoretically, and how we may find collections that achieve the theoretical minimum.

In 1974, Welch gave a family of lower bounds on c_{max} in terms of m and n .

Theorem 3.1 (Welch, 1974 [8]). *Let Z and c_{max} be defined as above. Then for each positive integer k , we have*

$$(c_{max})^{2k} \geq \frac{1}{m-1} \left(\frac{m}{\binom{n+k-1}{k}} - 1 \right). \quad (15)$$

Welch proved this theorem by bounding the average cross-correlation.

Lemma 3.2 (Welch). *Let $\{z_1, z_2, \dots, z_m\}$ be unit vectors in \mathbb{C}^n , then*

$$\frac{1}{m^2} \sum_{j=1}^m \sum_{k=1}^m |\langle z_j, z_k \rangle|^{2k} \geq \binom{n+k-1}{k}^{-1}. \quad (16)$$

By separating the diagonal terms from the sum and rearranging the summands, it is easy to see how (16) implies (15). Welch's original proof of (16) was combinatorial in nature. In 2012, Datta et al. provided a geometric proof based on examining the Gram matrix associated to Z and dimension counting. Both arguments are agnostic to whether the vectors are real or complex, and as far as we know, there is no attempt in the literature to investigate whether the bound can be improved if we further assume that the vectors are real.

Using the energy minimization property of rotationally invariant distributions, we are able to show that this is indeed the case.

Lemma 3.3. *Let $\{x_1, x_2, \dots, x_m\}$ be unit vectors in \mathbb{R}^n . Then*

$$\frac{1}{m^2} \sum_{j=1}^m \sum_{k=1}^m |\langle x_j, x_k \rangle|^{2k} \geq \frac{1 \cdot 3 \cdots (2k-1)}{n \cdot (n+2) \cdots (n+2k-2)}. \quad (17)$$

Remark 3.4. Since

$$\binom{n+k-1}{k}^{-1} = \frac{1 \cdot 2 \cdots k}{n \cdot (n+1) \cdots (n+k-1)},$$

we see that the new bound (17) is equal to the old one (16) for $k = 1$, and is strictly larger for $k > 1$.

Proof. Let X be uniformly distributed on the set $\{x_1, x_2, \dots, x_m\}$. Corollary 2.7 applies and we have

$$\mathbb{E} \langle X, X' \rangle^{2k} \geq \frac{1 \cdot 3 \cdots (2k-1)}{n \cdot (n+2) \cdots (n+2k-2)}$$

for any positive integer k . On the other hand, we also have

$$\mathbb{E} \langle X, X' \rangle^{2k} = \frac{1}{m^2} \sum_{j=1}^m \sum_{k=1}^m |\langle x_j, x_k \rangle|^{2k}. \quad \square$$

Let us illustrate the improved bound by revisiting an example from [5].

Example 3.5. Let x_1, x_2, \dots, x_7 be the columns of

$$\begin{bmatrix} 0.99 & 0.14 & 0.56 & -0.68 & 0.93 & -0.86 & 0.30 \\ 0.08 & 0.99 & 0.83 & 0.73 & -0.36 & -0.50 & 0.95 \end{bmatrix}.$$

This collection achieves the $k = 1$ Welch bound (16), and its energy²

$$\sum_{j=1}^7 \sum_{k=1}^7 |\langle x_j, x_k \rangle|^6 = 15.3128$$

was experimentally observed to be minimal over all collections of 7 unit vectors in \mathbb{R}^2 . However, the $k = 3$ Welch bound gives a lower bound of 12.25 for the energy, so there was a gap between theory and experiment. Using our improved bound (17), we get 15.3125, thereby bridging this gap completely.

Although the improved bounds do not hold for complex collections of vectors, we are nonetheless able to recover the original Welch bounds using the same circle of ideas and making a few adjustments.

Definition 3.6. For any random vector X in \mathbb{C}^n , let X_{uni} denote a random vector that is independent of X , has the same radial distribution as X , and whose distribution is invariant under unitary transformations. We call X_{uni} the *unitary symmetrization* of X .

With this definition, we can state the following complex version of Theorem 2.5.

Theorem 3.7. *Let X be a random vector in \mathbb{C}^n with finite moments of all orders. Then if X' is an independent copy of X , and X_{uni}, X'_{uni} are independent copies of its unitary symmetrization, we have*

$$\mathbb{E} |\langle X, X' \rangle|^{2k} \geq \mathbb{E} |\langle X_{uni}, X'_{uni} \rangle|^{2k} \quad (18)$$

for any positive integer k .

Proof. By considering the moment tensors

$$M_X^{2k} := \mathbb{E} X^{\otimes k} \otimes (X^*)^{\otimes k},$$

we may define a complex version of eccentricity tensors. Next, we replace $Q \sim \text{Haar}(O(n))$ with $U \sim \text{Haar}(U(n))$ in Lemma 2.4 to prove an orthogonality result analogous to (7). With this result, (18) follows immediately. \square

²To compute this value, we renormalized the vectors x_1, \dots, x_7 in order to reduce roundoff error.

We are now able to complete the proof of (16) with the help of the following version of Lemma 2.8.

Lemma 3.8 (Moments of complex spherical marginals). *Let θ be uniformly distributed on the complex sphere $S^{2n-1} \subset \mathbb{C}^n$. Then for any unit vector $v \in S^{2n-1}$ and any positive integer k , we have*

$$\mathbb{E}|\langle \theta, v \rangle|^{2k} = \binom{n+k-1}{k}^{-1}. \quad (19)$$

Proof. Let γ and g denote standard complex gaussians in 1 dimension and n dimensions respectively. Then

$$\mathbb{E}|\langle \theta, v \rangle|^{2k} = \frac{\mathbb{E}|\gamma|^{2k}}{\mathbb{E}\|g\|_2^{2k}}.$$

Since $|\gamma|$ is the norm of a two-dimensional standard real gaussian, while $\|g\|_2$ is the norm of a $2n$ -dimensional standard real gaussian, (19) follows from the calculations of gaussian integrals done in Lemma 2.8. \square

Remark 3.9. Given $Z = \{z_1, z_2, \dots, z_m\}$ a set of unit vectors in a Hilbert space \mathbb{H} , k a positive integer, define the set

$$Z^{(k)} = \{z_1^{\otimes k}, z_2^{\otimes k}, \dots, z_m^{\otimes k}\} \subset \text{Sym}^k(\mathbb{H}).$$

Datta et al.'s paper [5] characterized sets Z achieving equality in the k -th Welch average cross-correlation bound (16) as those for which $Z^{(k)}$ forms a tight frame for $\text{Sym}^k(\mathbb{H})$. Since our results show that this bound is not tight when \mathbb{H} is a real Hilbert space and $k > 1$, we have proved that tight frames of the form $Z^{(k)}$ do not exist for symmetric spaces of real tensors with $k > 1$. Indeed, this also holds true for generalized frames as defined by the same authors.

Remark 3.10. Datta et al. showed that the analogous statement for complex vector spaces is false. In fact, if θ is distributed uniformly on the complex sphere $S^{2n-1} \subset \mathbb{C}^n$, then

$$\mathbb{E}\theta^{\otimes k} \otimes (\theta^*)^{\otimes k} = \binom{n+k-1}{k}^{-1} I_{\text{Sym}^k(\mathbb{C}^n)}.$$

4. APPLICATIONS TO ENERGY OPTIMIZATION ON THE SPHERE

In a recent paper [2], Bilyk et al. presented a theorem characterizing probability measures minimizing geodesic distance energy integrals. This is an analogue of Björck's theorem from 1956 which characterized probability measures minimizing energy integrals based on Euclidean distance [3]. Björck proved his theorem by considering Riesz potentials, while Bilyk et al. proved their result using spherical harmonic expansions and the hemisphere Stolarsky principle. In this section, we show how to derive both results using the tensorization trick and the energy minimization property of the uniform distribution on the sphere.

Theorem 4.1 (Bilyk-Dai-Matzke, 2016). *For $\delta > 0$, define the geodesic energy integral*

$$G_\delta(\mu) := \int_{S^{n-1}} \int_{S^{n-1}} d(x, y)^\delta d\mu(x) d\mu(y), \quad (20)$$

where $d(x, y)$ denotes the geodesic distance between x and y . The maximizers of this energy integral over Borel probability measures on S^{n-1} can be characterized as follows:

- a) $0 < \delta < 1$: the unique maximizer of $G_\delta(\mu)$ is $\mu = \sigma$, the uniform measure.
- b) $\delta = 1$: $G_\delta(\mu)$ is maximized if and only if μ is centrally symmetric.
- c) $\delta > 1$: $G_\delta(\mu)$ is maximized if and only if $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$, i.e. the mass is supported equally by two antipodal points.

Proof. Observe that the geodesic distance $d(x, y)$ is simply the angle between x and y . As such, we have $d(x, y) = \arccos(\langle x, y \rangle)$. We may thus rewrite (20) as

$$G_\delta(\mu) = \mathbb{E} \arccos(\langle X, X' \rangle)^\delta,$$

where X and X' are independent random vectors with distribution μ .

Let us start by proving part b). It is an exercise to show that the even derivatives of \arccos vanish at 0, while the odd derivatives are strictly negative at 0. For $-1 < t < 1$ may hence write \arccos as its Taylor series

$$\arccos(t) = \frac{\pi}{2} - \sum_{k=0}^{\infty} a_{2k+1} t^{2k+1} \quad (21)$$

where $a_{2k+1} > 0$ for all k . We claim that in fact, the above formula holds for all t in the *closed* interval $[-1, 1]$, and furthermore that the series is absolutely convergent. This is the content of Lemma 4.2 to come. As a result, we may use Fubini to interchange sums and expectations, thereby writing

$$\mathbb{E} \arccos(\langle X, X' \rangle) = \frac{\pi}{2} - \sum_{k=0}^{\infty} a_{2k+1} \mathbb{E} \langle X, X' \rangle^{2k+1}.$$

Since $\mathbb{E} \langle X, X' \rangle^{2k+1} \geq 0$ for each k by identity (4), this last expression is maximized if and only if $\mathbb{E} \langle X, X' \rangle^{2k+1} = 0$ for every non-negative integer k . By the same identity, this happens if and only if all odd moments of X are zero, i.e. if and only if X is centrally symmetric. This proves the case $\delta = 1$.

Now let $0 < \delta < 1$. We claim that for $-1 \leq t \leq 1$, we may write

$$\arccos(t)^\delta = \left(\frac{\pi}{2}\right)^\delta - \sum_{k=1}^{\infty} a_k t^k \quad (22)$$

where $a_k > 0$ for all $k > 0$, and that the series is absolutely convergent. Lemma 4.3 (to come) tells us that the Taylor series of $\arccos(t)^\delta$ has this form, which combined with Lemma 4.2 proves this claim. As such, we may again use Fubini to write

$$\mathbb{E} \arccos(\langle X, X' \rangle)^\delta = \left(\frac{\pi}{2}\right)^\delta - \sum_{k=1}^{\infty} a_k \mathbb{E} \langle X, X' \rangle^k. \quad (23)$$

By identity (4), $\mathbb{E} \langle X, X' \rangle^k \geq 0$ for any distribution, while by Corollary 2.7, the uniform measure uniquely minimizes all of these moments simultaneously. As such, we see that it is the unique maximizer of $G_\delta(\mu)$.

The remaining case where $\delta > 1$ is easy and does not require a proof using our methods. For completeness, we repeat the proof given by the original authors [2]. Since $d(x, y) \leq \frac{\pi}{2}$, we have

$$G_\delta(\mu) \leq \left(\frac{\pi}{2}\right)^{\delta-1} \int_{S^{n-1}} \int_{S^{n-1}} d(x, y) d\mu(x) d\mu(y) \leq \left(\frac{\pi}{2}\right)^\delta.$$

The first inequality is tight whenever $d(x, y)$ only takes the values $\frac{\pi}{2}$ and 0, while by part b), the second inequality becomes equality when μ is centrally symmetric. Together, these imply that $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$ for some $p \in S^{n-1}$. \square

Lemma 4.2. *Let f be a function that is continuous on $[-1, 1]$ and that agrees with its Taylor series at 0 on the open interval $(-1, 1)$. Suppose further that all but finitely many of its derivatives at 0 have the same sign. Then the series is absolutely convergent over the closed interval $[-1, 1]$, and agrees with f over the interval.*

Proof. By subtracting off polynomials and negating the function if necessary, we may assume without loss of generality that the Taylor series for $f(t)$ is given by $\sum_{k=0}^{\infty} c_k t^k$ where $c_k \geq 0$ for all k . By the monotone convergence theorem, together with our assumptions on f , we have

$$\sum_{k=0}^{\infty} c_k = \lim_{t \rightarrow 1^-} \sum_{k=0}^{\infty} c_k t^k = \lim_{t \rightarrow 1^-} f(t) = f(1).$$

As such, the series $\sum_{k=0}^{\infty} c_k$ is absolutely convergent, and the Taylor series is also absolutely convergent on the closed interval $[-1, 1]$. Finally, we can apply the dominated convergence theorem to see that $f(-1) = \sum_{k=0}^{\infty} c_k (-1)^k$. \square

Lemma 4.3. *Let f be a function that has derivatives of all orders at 0 and let $0 < \alpha < 1$. Suppose $f(0) > 0$ and $f'(0) < 0$, while all higher derivatives $f^{(k)}$ at 0 are non-positive, then all derivatives of f^α at 0 are strictly negative.*

Proof. Let $F(t) = f(t)^\alpha$. By induction, one may observe that for any positive integer k , $F^{(k)}(t)$ is a sum of 2^{k-1} terms of the form

$$g_{\vec{n}}(t) := f(t)^{\alpha-j} \left(\prod_{i=0}^{j-1} (\alpha - i) \right) \left(\prod_{i=0}^{j-1} f^{(n_i)}(t) \right),$$

where $1 \leq j \leq k$, and $\vec{n} = (n_0, n_1, \dots, n_{j-1})$ is a vector of positive integers. If there is some index i such that $f^{(n_i)}(0) = 0$, then $g_{\vec{n}}(0) = 0$. Otherwise, $\prod_{i=0}^{j-1} f^{(n_i)}(0)$ is a product of j negative numbers and so has sign $(-1)^j$. On the other hand, our assumption on α imply that $\left(\prod_{i=0}^{j-1} (\alpha - i) \right)$ is a product of one positive number and $j-1$ negative numbers, and so has sign $(-1)^{j-1}$. As such, $g_{\vec{n}}(0) \leq 0$.

Finally, notice that $F^{(k)}(0)$ always contains the term

$$g_{(1,1,\dots,1)}(0) = f(t)^{\alpha-k} \left(\prod_{i=0}^{k-1} (\alpha - i) \right) f'(0)^k.$$

Since we have assumed that $f'(0) < 0$, this term is strictly negative. As such, $F^{(k)}(0)$ is also negative, as was to be shown. \square

In the course of proving the previous theorem, we have in fact proved the following more general result.

Theorem 4.4. *Let F be a function on $[-1, 1]$ that is given by the power series*

$$F(t) = a_0 - \sum_{k=1}^{\infty} a_k t^k, \quad (24)$$

where $a_k \geq 0$ for all $k > 0$. Then the energy integral

$$I_F(\mu) := \int_{S^{n-1}} \int_{S^{n-1}} F(\langle x, y \rangle) d\mu(x) d\mu(y) \quad (25)$$

is maximized over all Borel probability measures on S^{n-1} by the uniform measure. Furthermore, if $a_k > 0$ for all $k > 0$, then the maximizer is unique.

Let us see how we may apply this more general theorem to recover Björck's original result.

Theorem 4.5 (Björck, 1956). *For $\delta > 0$, define the energy integral*

$$E_\delta(\mu) = \int_{S^{n-1}} \int_{S^{n-1}} \|x - y\|^\delta d\mu(x) d\mu(y). \quad (26)$$

The maximizers of this energy integral over Borel probability measures on S^{n-1} can be characterized as follows:

- (1) $0 < \delta < 2$: the unique maximizer of $E_\delta(\mu)$ is $\mu = \sigma$, the uniform measure.
- (2) $\delta = 2$: $E_\delta(\mu)$ is maximized if and only if the center of mass of μ is at the origin.
- (3) $\delta > 2$: $E_\delta(\mu)$ is maximized if and only if $\mu = \frac{1}{2}(\delta_p + \delta_{-p})$, i.e. the mass is supported equally by two antipodal points.

Proof. We rewrite (26) as

$$E_\delta(\mu) = \mathbb{E} \|X - X'\|_2^\delta$$

where X and X' are independent random vectors with distribution μ . The easy case $\delta > 2$ is proved exactly as in Theorem 4.1. The case $\delta = 2$ is also clear, for we may write $\|X - X'\|^2 = 2 - 2\langle X, X' \rangle$, and by identity (4),

$$E_2(\mu) = 2 - \mathbb{E} \langle X, X' \rangle = 2 - \|\mathbb{E} X\|_2^2.$$

This is maximized if and only if $\mathbb{E} X = 0$.

For $0 < \delta < 2$, we set $f(t) = 2 - 2t$ and $F(t) = f(t)^{\delta/2}$. Then f and $\alpha = \delta/2$ satisfy the hypotheses of Lemma 4.3, so $F^{(k)}(0) < 0$ for all positive integers k . This, together with Lemma 4.2, implies that F satisfies the hypothesis of Theorem 4.4. Since

$$E_\delta(\mu) = \int_{S^{n-1}} \int_{S^{n-1}} (2 - 2\langle x, y \rangle)^{\delta/2} d\mu(x) d\mu(y) = \int_{S^{n-1}} \int_{S^{n-1}} F(\langle x, y \rangle) d\mu(x) d\mu(y),$$

we can conclude that $E_\delta(\mu)$ is uniquely maximized by the uniform measure. \square

Remark 4.6. In their paper [2], Bilyk et al. remarked that while the Euclidean and geodesic distances are both metrics on the sphere, the phase transition for the behavior of their energy integrals is different. In the Euclidean case, Björck's theorem shows that it occurs at $\delta = 2$, while in the geodesic case, Bilyk et al.'s theorem shows that it occurs at $\delta = 1$. This peculiar phenomenon is explained by our unified proof of both results.

In both cases, the existence of a phase transition as we let δ decrease to 0 is asserted by Lemma 4.3 and Theorem 4.4. If the integrand satisfies the hypotheses of Lemma 4.3 for some δ_0 , then for all $0 < \delta < \delta_0$, the integrand will satisfy the hypothesis of Theorem 4.4, from which we can conclude that the unique maximizer is the uniform measure. For the Euclidean integral, we have $\delta_0 = 2$, while for the geodesic integral, we have $\delta_0 = 1$.

Remark 4.7. Bilyk et al. were also interested in understanding continuous functions F for which the uniform measure σ is the unique minimizer of I_F as defined in (25). They managed to characterize these functions as those for which all non-constant Gegenbauer coefficients are strictly positive, i.e.

$$\hat{F}(k, \lambda) > 0$$

for all positive integers k , and where $\lambda = \frac{n}{2} - 1$. On the other hand, by flipping signs, Theorem 4.4 implies that a sufficient condition for this to happen is to require all non-constant Taylor series coefficients to be strictly positive.

ACKNOWLEDGEMENTS

I would like to thank Dmitriy Bilyk, Han Huang, and Roman Vershynin for insightful discussions on these topics.

REFERENCES

- [1] BILLINGSLEY, P. *Probability and Measure*. Wiley, 1986.
- [2] BILYK, D., DAI, F., AND MATZKE, R. *Stolarsky principle and energy optimization on the sphere*. Preprint. Available at <https://arxiv.org/abs/1611.04420>.
- [3] BJÖRCK, G. *Distributions of positive mass, which maximize a certain generalized energy integral*. Arkiv för Matematik 3, 255-269, 1956.
- [4] ÇINLAR, E. *Probability and Stochastics*. Graduate Texts in Mathematics, Springer, 2011.
- [5] DATTA, S., HOWARD, S., AND COCHRAN, D. *Geometry of the Welch bounds*. Linear Algebra and its Applications 437, 2455-2470, 2012.
- [6] DONOHO, D., AND ELAD, M. *Optimally sparse representation in general (nonorthogonal) dictionaries via l_1 minimization*. PNAS 100, 5, 2197-2202, 2003.
- [7] VERSHYNIN, R. *Introduction to the non-asymptotic analysis of random matrices*. Compressed Sensing, 210-268, Cambridge University Press, Cambridge, 2012.
- [8] WELCH, L. *Lower bounds on the maximum cross correlation of signals*. IEEE Transactions on Information Theory 20, 3, 397-399, 1974.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF MICHIGAN, 530 CHURCH STREET, ANN ARBOR, MI 48109
E-mail address: yanshuo@umich.edu